
Article:
Shimizu, Seishi orcid.org/0000-0002-7853-1683, Abbott, Steven, Adamska, Katarzyna et al. (1 more author) (2019) Quantifying non-specific interactions via liquid chromatography. Analyst. pp. 1632-1641. ISSN 0003-2654

https://doi.org/10.1039/C8AN02244E

Quantifying non-specific interactions via liquid chromatography

Seishi Shimizu,[1,] Steven Abbott,[2,3] Katarzyna Adamska,[4] Adam Voelkel[4]*

[1]York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, United Kingdom.

[2]Steven Abbott TCNF Ltd., 7 Elsmere Road, Ipswich, Suffolk IP1 3SZ, United Kingdom.

[3]School of Mechanical Engineering, University of Leeds, Leeds LS2 9JT, United Kingdom.

[4]Institute of Chemical Technology and Engineering, Poznan University of Technology, Berdychowo 4, 60-965 Poznan, Poland.

AUTHOR INFORMATION

Corresponding Author:

Seishi Shimizu

York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, United Kingdom

Tel: +44 1904 328281, Fax: +44 1904 328281, Email: seishi.shimizu@york.ac.uk

ABBREVIATIONS

KB, Kirkwood-Buff; KBI, Kirkwood-Buff integral; CMC, Critical micelle concentration

ABSTRACT

Determinations of solute-cosolute interactions from chromatography have often resulted in problems, such as the "antibinding" (or a negative binding constant) between the solute and micelle in micellar liquid chromatography (MLC) or indeterminacy of salt-ligand binding strength in high-performance affinity chromatography (HPAC). This shows that the stoichiometric binding models adopted in many chromatographic analyses cannot capture the non-specific nature of solvation interactions. In contrast, an approach using statistical thermodynamics handles these complexities without such problems and directly links chromatographic data to, for example, solubility data via a universal framework based on Kirkwood-Buff integrals (KBI) of the radial distribution functions. The chromatographic measurements can now be interpreted within this universal theoretical framework that has been used to rationalize small solute solubility, biomolecular stability, binding, aggregation and gelation. In particular, KBI analysis identifies key solute-cosolute interactions, including excluded volume effects. We present (i) how KBI can be obtained directly from the cosolute concentration dependence of the distribution coefficient, (ii) how the classical binding model, when used solely as a fitting model, can yield the KBIs directly from the literature data, and (iii) how chromatography and solubility measurements can be compared in the unified theoretical framework provided via KBIs without any arbitrary assumptions about the stationary phase. To perform our own analyses on multiple datasets we have used an "app". To aid readers'

understanding and to allow analyses of their own datasets, the app is provided with many datasets and is freely available on-line as an open-source resource.

1. Introduction

The way that solutes interact in the aqueous environment can be modulated through the addition of small molecules ("cosolutes"), leading to solubility enhancement,[1–4] stabilizing biopolymer folding, controlling protein-ligand interaction and controlling protein aggregation and gelation.[5–8] To this end, many different cosolutes (salts, small organic molecules, hydrotropes, micelles and polymers) have been used. These can exhibit different degrees of self-aggregation (micellar surfactants, hydrotropes and cosolvents) and different modes of interaction specificity with the solute (e.g. encapsulation by cyclodextrins, incorporation into micelles, electrostatic interactions with salts, and weak, non-specific interactions with osmolytes).[9–11]

Chromatography is a powerful technique to quantify solute-cosolute interaction due to its superiority in speed and simplicity compared to thermodynamic techniques, which are often slow and painstaking.[5,12] Different chromatographic approaches, targeting a particular range of self-aggregation and specificity, have been developed, including

(i) micellar liquid chromatography (MLC) to quantify micelle-solute affinities;[13–17]

(ii) high performance affinity chromatography (HPAC) to measure protein-ligand affinities[18–26]

However, the interpretations of the data from these techniques are still based on ideas of stoichiometric solute-cosolute binding,[13–20] which have not only been invalidated and superseded

in the recent development in complex solutions, but have also been shown to contradict the actual chromatographic data.

Indeed, "antibinding" or negative solute-cosolute binding constants have been reported in MLC for decades,[27–29] which is contradictory in principle to the basic assumption, i.e., the stoichiometric binding between the solute and cosolute. Moreover, the salt-ligand binding constant, one of the fundamental parameters for HPAC, sometimes turned out to be indeterminate[18] by experimental data; several different binding constants may also be required for some systems.[18] Analysing chromatographic data by stoichiometric binding models has thus led to fundamental difficulties.

Such fundamental difficulties cannot be resolved by data acquisition. What is needed instead is a reconsideration of the theoretical foundation upon which the experimental data are analysed. Here we show that the stoichiometric binding models can be replaced by a general molecular (statistical) thermodynamic theory, which enables the quantification of non-specific interactions based on the realistic picture of such interactions in solutions.[3,4] Instead of the binding constant which, by definition, can only be positive (and where rectifying this has led to more complex assumptions),[30–32] we advocate the use of the Kirkwood-Buff integrals (KBIs),[33–35] defined from the distribution of molecules around the solute, for the following advantages:

- Non-specific interactions are defined directly from a fundamental property of solution structure, i.e., molecular distribution functions[3,4]
- Association and exclusion of the cosolutes around the solute can both be treated[36–38]
- The analysis does not depend on any presumed linearity of any plot[39]

In addition, characterising the non-specific and non-stoichiometric nature of solute-cosolute interactions by KBIs has resolved decades-long confusions and controversies in other scientific disciplines which had previously used stoichiometric modelling of non-stoichiometric interactions.[8,10,36,37] There, as here, the key for resolution was to abandon the stoichiometric models and to replace them with molecular distribution functions.[3,4,8] Despite the clarity attained in other fields of research, chromatographic analysis on cosolute effects still suffers from the lack of clarity. We will demonstrate that analysing chromatographic data based on the molecular distribution functions will lead to a novel, universal method that do not suffer from the fundamental difficulties of the traditional methods.[18,27–29]

Hence the goal of this paper is fourfold:

1. To establish a universal theory of cosolvent effects in chromatography regardless of the degree of cosolvent self-association, solute-cosolvent binding strength and specificity.
2. To link the previous theoretical models to the universal theory so that all the fitting parameters reported in the literature can immediately be useful in quantifying the cosolvent effect on a molecular basis in the framework of the universal theory.
3. To clarify the presumed relationship between chromatography and solubility for MLC in terms of KBIs, thereby providing the criterion to judge whether chromatography can facilitate high-throughput determination of solute-micelle affinities.
4. To highlight the problem of the necessary assumption made by stoichiometric models that the cosolute does not interact with the stationary phase.

Thus, the main focus of this paper is to propose a new approach to chromatographic data analysis and interpretation, in order to overcome the difficulties and paradoxes arising from the traditional

analysis methods. As in our previous papers,[3,39,40] the calculations, based directly on our novel approach, throughout have been performed using an "app", which is freely available on-line, along with all the datasets used, so that readers can check the approach for themselves and load and analyse their own datasets.

2. A statistical thermodynamic foundation for chromatography

2.1 High Performance Affinity Chromatography (HPAC)

We consider the mobile phase comprised of dilute ligand ($i = u$), water ($i = 1$) and cosolute ($i = 2$) molecule. Proteins, that bind ligand in a specific manner, have been fixed onto the stationary phase. Let us denote the concentration of the species c_i. At a given c_2, the number of receptor active sites n, the volume of the mobile phase V_m, and the retention factor k can all be measured, from which the distribution constant K of the solute can be calculated using a well-known formula as

$$k = K\frac{n}{V_m} \tag{1}$$

The distribution coefficient K, under the instantaneous equilibrium assumption, signifies ligand partitioning between the mobile and stationary phases, as

$$K = \frac{c_u^s}{c_u} \tag{2}$$

where c_u^s is the ligand concentration in the stationary phase. Consequently, K can be linked via statistical thermodynamics to the free energy $\Delta\mu_u^*$ of transferring a ligand from a fixed centre of mass position in the stationary phase to that in the mobile phase as

$$\Delta\mu_u^* = -RT \ln \frac{c_u}{c_u^s} = RT \ln K$$

(3)

With the above setup, chromatographic measurements can now be analysed in the framework of statistical thermodynamics. To do so, the distribution constant K of ligands is measured along with the concentration of non-micellar cosolutes, c_2. Then a rigorous statistical thermodynamic result[7,36,37] can be used to interpret the distribution constant when the cosolutes are dilute

$$-\frac{1}{RT}\left(\frac{\partial \Delta\mu_u^*}{\partial c_2}\right)_{T,P,c_u\to 0} = -\left(\frac{\partial \ln K}{\partial c_2}\right)_{T,P,c_u\to 0} = \Delta G_{u2} - \Delta G_{u1}$$

(4)

in terms of the change Δ of the Kirkwood-Buff integral (KBI) that accompanies ligand dissociation. The KBI between the solute and the species i at state α, either "b" for bound ligand-protein pair or "d" for dissociated, is defined as

$$G_{ui}^{\alpha} = 4\pi \int dr\, r^2\, [g_{ui}(r) - 1]$$

(5)

in which $g_{ui}(r)$ is the radial distribution function between the solute and the species i.[33–35] Hence ΔG_{ui} is defined as the difference in KBI between the dissociated and bound states, i.e., $\Delta G_{ui} = G_{ui}^d - G_{ui}^b$.[36,37] Such a KBI difference can be evaluated from how the retention factor k depends on cosolute concentration c_2, because the c_2 dependence of K in Eq. (4) comes from the c_2 dependence of k.

It is natural to consider that water and cosolute both interact not only with the ligand but also with the protein, hence the chromatographic distribution coefficient K should correspond to the ligand-protein dissociation constant in the thermodynamics of protein-ligand binding. This is

indeed in line with Eq. (4) in which the cosolute effect on protein-ligand binding is driven by the *differences* of KBIs. However, earlier literature on the cosolute effect considered the mutual binding of cosolutes (or "displacers") and ligands to a common site on the stationary phase,[41–43] which does not directly consider cosolute-ligand interaction in the mobile phase. The thermodynamic binding model was later extended to rectify this shortcoming, only to result in indeterminate binding constant (see Section 3.1).[18]

Thus, we advocate that KBIs be used as the universal measure to quantify affinities from chromatographic measurements, instead of a number of different models and conventions developed for particular applications.[13–20] The universality of KBIs comes from its definition as net affinity, based directly on solution structure.[33–35] To appreciate its meaning, let us first note that solute-cosolute distribution function, $g_{u2}(r)$, tends to 1 far away from the solute (at large r), where the solution structure is no longer affected by the presence of the solute and therefore is the same as the bulk solution. The attractive regions (i.e., r with $g_{u2}(r) > 1$) contributes positively to KBIs whereas repulsion (i.e., r with $g_{u2}(r) < 1$) contributes negatively. In summary, the KBI (Eq. (5)) signifies the net increase in the concentration of species i around the solute compared to the bulk solution.[3,39]

Thus, the increment of protein-cosolute and protein-water interactions that accompany ligand dissociation, ΔG_{u2} and ΔG_{u1}, competitively contribute to drive the ligand dissociation. To identify which of the two is the dominant contribution, Eq. (4) from chromatographic should be combined with the partial molar volume change Δv_u^0 that accompanies protein-ligand dissociation, which is known to yield ΔG_{u1} via[7,36,37]

$$\Delta G_{u1} = -\Delta v_u^0 \quad (6)$$

Note that Eqs. (4) and (6) have been derived under dilute cosolute concentration as has been commonly performed in chromatographic measurements, and can be generalized straightforwardly to higher cosolute concentrations.[7,36,37]

To extract KBI from experimental data of retention factor, k, versus cosolute concentration, the following steps are required:

(i) fit the data to a convenient function,

(ii) use the analytical derivative of that function at any concentration to calculate ΔG_{u2} from Eq. (4) as ΔG_{u1} is negligibly small.

Indeed, evidence in the literature[7,8,36,37] shows that that ΔG_{u1} is usually in the order of 10^1-10^2 $\mathrm{cm}^3\ \mathrm{mol}^{-1}$, which, as can be seen from the app, is several orders of magnitude smaller than ΔG_{u2}. Such an order-of-magnitude analysis, which has led to a drastic simplification in data analysis, was made possible via a novel link between the chromatographic k and KBIs, which has made it possible to be compared with the volumetric data on binding. Such a negligibility of ΔG_{u1} simplifies the application of Eq. (4) to salts as cosolutes. ΔG_{u2} per salt has been calculated by taking c_2 as the concentration of salts whereas the per-ion ΔG_{u2}, commonly employed in the literature of KBI calculation, can be obtained simply multiplying per-salt ΔG_{u2} by the number of ions per salt.[38,44,45]

The process is conveniently done via the app shown in Figure 1 which can perform the analysis of data from a wide range of different systems assembled and placed into a uniform format by the Peyrin group,[18] though, of course, our analysis is different from theirs. We emphasise that the app,

which already contains at least 4 different classes of data sets, is provided here to enable the readers to test this new, universal approach. The data used in the app is necessarily of limited quality because it has been extracted from the graphs and converted from the log-log format of the original. Users can also load their own data in the simple format specified in the app. To fit the wide variety of data, a pragmatic choice of fitting equation was made that is generally robust and does a reasonable job with the 11 examples extracted from the paper.

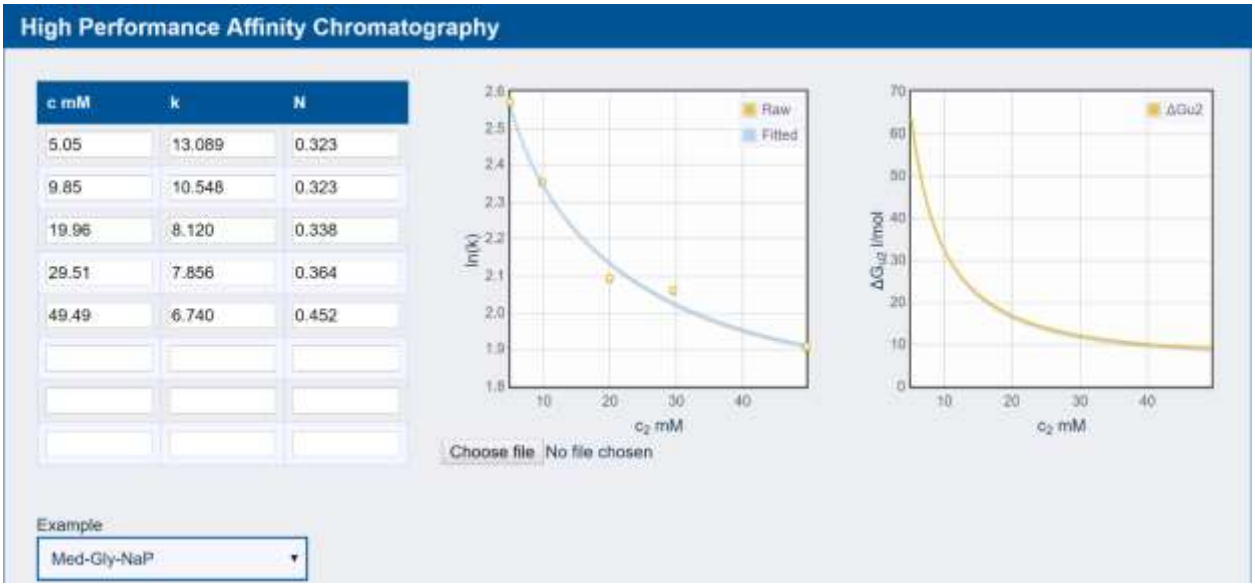

Figure 1. An interactive statistical thermodynamic analysis of High Performance Affinity Chromatography (HPAC) data taken from Slama et al.[18] for medetomide-R-1-glycoprotein-sodium phosphate in the presence of salts. The raw data are shown in the table in the app. The $\ln k$ versus salt concentration (c_2) data are fitted to a convenient function and the solute-cosolute KBI (ΔG_{u2}) values are extracted from the analytical derivative of the curve. Per-salt ΔG_{u2} has been reported for salt. The app is available at https://www.stevenabbott.co.uk/practical-chromatography/HPAC.php

2.2. Micellar Liquid Chromatography (MLC)

The universality of the statistical thermodynamic foundation presented in Section 2.1 can be demonstrated by seeing how easy it is to adapt it to a different chromatographic technique, specializing exclusively in micellar cosolutes, which previously had been analyzed using a separate theoretical model. MLC deals with micellar cosolutes, for which we consider surfactant monomer ($i = 2$) and surfactant micelle ($i = 3$), whose aggregation number is m,[4,9] under micelle-monomer equilibrium above the critical micelle concentration (CMC). Due to the micelle-monomer equilibrium, the solvent mixture (water, monomer and micelle) still behaves like an effective two-component mixture; indeed, chromatographic measurements are conducted in the concentrations much larger than CMC where surfactants are predominantly in a micellar form.[4,9] In addition to the concentration c_i defined in Section 2, we also use c_2' as the total concentration of surfactant monomer, such that $c_2' = mc_3$ for the case of predominant micelle formation.[4,9]

In MLC, the solute's distribution coefficient K is measured along with the surfactant concentration, c_2', which can be linked to KBIs via a rigorous statistical thermodynamic result derived under the condition that the surfactants are predominantly in the micellar form,[9] as

$$-\frac{1}{RT}\left(\frac{\partial \Delta\mu_u^*}{\partial c_3}\right)_{T,P,c_u\to 0} = \Delta G_{u3} - \Delta G_{u1} \tag{7}$$

where ΔG_{ui} signifies the difference of KBI between the mobile ($\alpha = m$) and stationary ($\alpha = s$) phases, i.e., $\Delta G_{ui} = G_{ui}^m - G_{ui}^s$.

Note, according to the implicit assumption of MLC, that the surfactant predominantly affects the mobile phase, so K, according to MLC, is the measure of solute affinity to the mobile phase, independent of surfactant concentration. If this is true, then the right-hand side of Eq. (7) actually

involves the mobile phase only (i.e., $G_{u3}^{m} - G_{u1}^{m}$). This means that it can also be determined from how solubility (measured by classical techniques) depends on surfactant concentration. If this chain of logic is true then there is a common foundation for solubility and chromatographic measurements, leading to an expectation that micellar solubilization can be measured much more efficiently using chromatography than from solubility measurements. In Section 3.3., the comparison between some MLC-based and solubility-based values is made. The evidence suggests that this fast-track method is not currently reliable.

Eq. (7) shows that micellar solubilization is due to the stronger solute-cosolute interaction increase (upon stationary to mobile transfer of a solute) compared to solute-water.[9] In quantifying micellar solubilization, the accumulating evidence for the multiple possible locations (from interior to the surface) for a solute in the micellar system[46–49] has posed much difficulties. However, Eq. (7) is valid regardless of the solute location.[4,9] If the solute is buried inside, then $g_{u3}(r)$ has a sharp peak near $r = 0$. If the solute is bound between the hydrophobic chains of the surfactant molecules, then $g_{u3}(r)$ peak shifts to a larger *r* but still less than the micelle radius. If the solute is bound on the surface, a sharp $g_{u3}(r)$ peak is observed at the micelle-solute contact distance. Wherever the solute is located, KBI can link solute-micelle affinity to solubilization.

In the current analysis of experimental data, surfactant concentration is commonly used instead of the micellar concentration. The use of total surfactant concentration c_2', with the use of the aggregation number m yields the expression conforming to this practice[4,9]

$$-\left(\frac{\partial \ln K}{\partial c_2'}\right)_{T,P,c_u \to 0} = \frac{1}{m}(\Delta G_{u3} - \Delta G_{u1}) \qquad (8)$$

Eq. (8) clarifies the two competing driving forces at work for solubilization.

(i) The large m, which refers to the effective reduction in the number of hydrotropes, hence an apparent inefficiency due to micellization.[9]

(ii) The large ΔG_{u3} more than compensates for this factor, making many micellar hydrotropes effective solubilizers.[9]

Note, due to the existence of monomer-micelle equilibrium, that solute-micelle and solute-monomer KB integrals cannot be determined independently.[4,11] However, solute-micelle and solute-water KBIs can both be determined. To do so, just like HPAC, the change of partial molar volume v_u^0 of the solute accompanying the transfer from stationary to mobile phases, should be used in conjunction,[4,9,36] as

$$\Delta G_{u1} = -\Delta v_u^0 \tag{9}$$

To extract KBI from experimental data of retention factor, k, versus concentration of surfactant just two steps are required: fit the data (ln K versus c_2') to a convenient polynomial, then use the analytical derivative of that polynomial at any concentration to calculate ΔG_{u3} from Eq. (8), using any convenient estimate of ΔG_{u1}. Although K requires n/V_m, because this appears only as a small constant (e.g. ln(100)~5), it can be ignored in calculation of the relatively large ΔG_{u3} values. Similarly, the ΔG_{u1} term is usually negligible[9,36] and is not required for the app. The process is conveniently done via the app shown in Figure 2 which is performing the analysis of some classic data from Armstrong and Stine. The user can select one of the many datasets we have analysed, or users can load their own datasets if provided in a simple format described in the app. Moreover, readers are encouraged to use this free application to check the validity of new approach using their own experimental data.

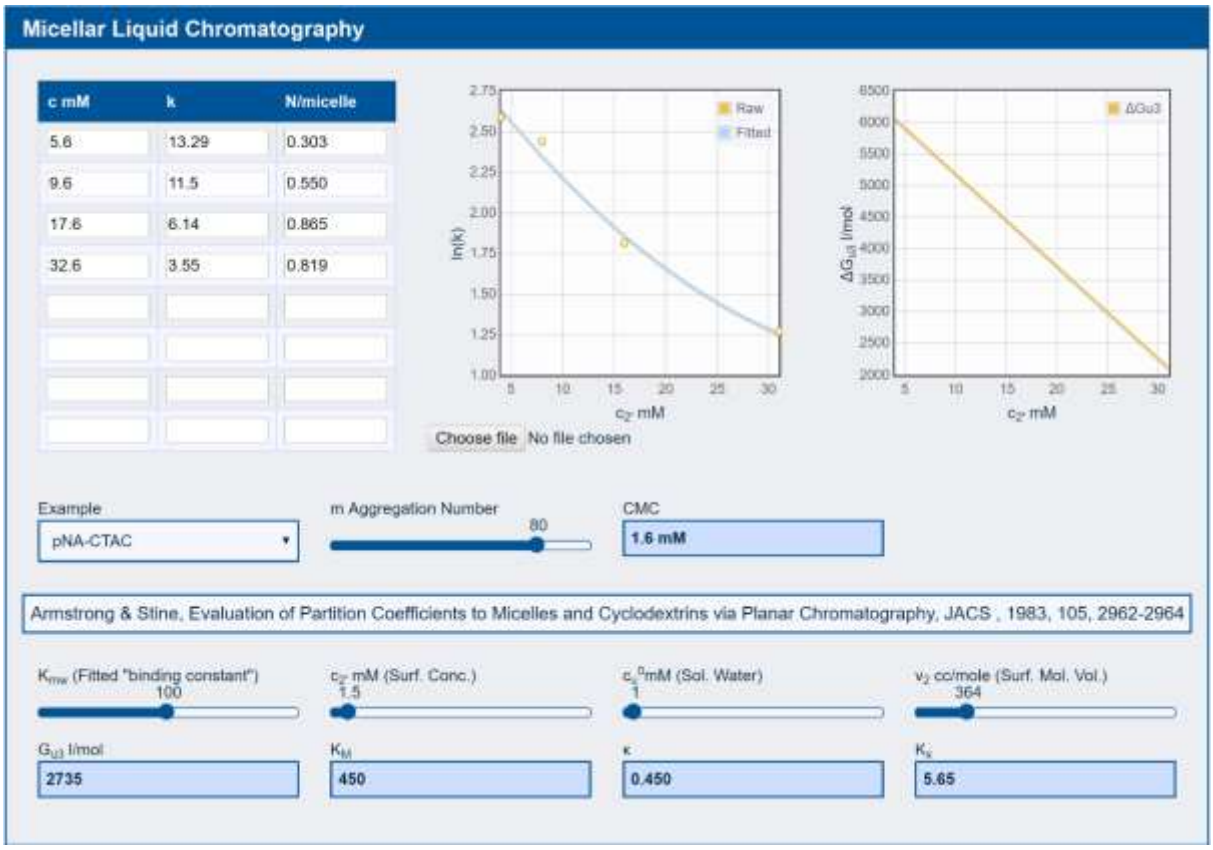

Figure 2. A statistical thermodynamic analysis of Micellar Liquid Chromatography (MLC) data, taken from literature,[13,29,50–52] which interactively shows the route from raw data to fitting to calculation of solute-micelle KBI (ΔG_{u3}) values. The sliders allow users to change the estimates of potentially uncertain values such as aggregation number. The significance of the lower part of the app is discussed below. The app is available at https://www.stevenabbott.co.uk/practical-chromatography/MLC.php

3. Statistical thermodynamics versus stoichiometric binding models

3.1. Breakdown of stoichiometric binding in HPAC

As demonstrated in Section 2, KBIs, the universal affinity measure, can be calculated straightforwardly from experimental data on the dependence of retention factor k on cosolute concentration. Such a clarity, generality and universality cannot be obtained from the traditional analysis based on thermodynamic models, as will be demonstrated below.

Prior to the statistical thermodynamic theory, the effect of cosolutes on solvation had long been modelled using the competitive solvent binding model.[31,32,53–56] Here we do not discuss many fundamental difficulties and confusions that necessitated the replacement of this model by statistical thermodynamics, for which our reviews may be consulted.[3,4,8]

We focus instead on the difficulties encountered by the previous analysis of cosolute (salt) effect on protein-ligand binding conducted via the competitive solvent binding model.[18–20] According to this model, the number of cosolutes released on protein-ligand dissociation, Δn_2, as well as the average binding constant of each cosolute κ, are the key to parameters to be quantified from experiments, via[18,54,55]

$$\ln k = \ln k_0 + \Delta n_2 \ln(1 + \kappa c_2) \tag{10}$$

where k_0 is a constant. In the analysis, however, a linear dependence of $\ln k$ on $\ln c_2$ has been reported for the ligand-protein combinations of aspartame and α-chymotrypsin, medetomide and α-1-glycoprotein, and beraprost and α-1-glycoprotein,[18] that have led to the determination of Δn_2 via

$$\frac{\partial \ln k}{\partial \ln c_2} \simeq \Delta n_2 \tag{11}$$

Note that Eq. (11) is the $\kappa \to \infty$ limit of Eq. (10). This relationship, however, is problematic for the following reasons: (i) κ, another key parameter for the model, is rendered undeterminable; (ii) Eqs. (4) and (11) leads to the KBI,

$$\Delta G_{u2} - \Delta G_{u1} = -\frac{\partial \ln k}{\partial c_2} = -\frac{\partial \ln k}{\partial \ln c_2}\frac{\partial \ln c_2}{\partial c_2} = -\frac{\Delta n_2}{c_2} \quad (12)$$

which diverges at $c_2 \to 0$, which is unphysical, showing the difficulty with regards to this linear relationship between $\ln k$ and $\ln c_2$. Moreover, highly non-linear relationships, whenever observed, have been fitted by assuming multiple strengths of salt binding constants, which complicates the analysis with arbitrary assumptions.[18]

In contrast, as has been demonstrated in Section 2, the calculation of KBIs is so much simpler and universal. Firstly, unlike the KBI approach which uses the same model, the thermodynamic model for salts as cosolutes adopted by HPAC[18–20,54,55] is completely different from that for surfactant cosolutes developed for MLC (Section 3.2). Secondly, although calculation of KBIs still requires fitting functions, unlike the traditional binding models which require the data to be forced into linear plots (such as the Hill or Klotz plot[57]), the fitting functions are solely for the purpose of calculating the gradient of $\ln k$ with respect to c_2. Thirdly, KBIs can be used regardless of cosolute self-association.[4,9]

3.2. Obtaining KBIs from existing MLC binding model values

There is a further advantage to the KBI approach: we can convert, with obvious limitations, previously published data into KBI values enabling the wealth of data in the literature to be used within the KBI context.

The idea is to use binding model data as an empirical[28] fitting model for experimental data. This enables us to obtain KBIs directly from the results in the literature that have relied upon the stoichiometric model. To do so, let us start from the distribution/partition coefficient K, and rewrite it in the language of the binding coefficients of Armstrong and Stine,[28] while using the molarity concentration throughout, as

$$K = \frac{c_u^s}{c_u} = \frac{c_u^s}{c_u^{aq} + c_u^m} = \frac{\frac{c_u^s}{c_u^{aq}}}{\frac{c_u^m}{c_u^{aq}} + 1} \tag{13}$$

where c_u^{aq} and c_u^m represent the solute concentrations in the "aqueous" (bulk water) (sub)phase and in the micelle, respectively. Following Armstrong and Stine, Eq. (13) can be rewritten using the volume fraction of the micelle, $c_3 v_3$ (where v_3 is the partial molar volume of the micelle), as well as the solute's micelle-water and stationary-water partition constants, K_{mw} and K_{sw}, as[28]

$$K_{mw} = \frac{1 - c_3 v_3}{c_3 v_3} \frac{c_u^m}{c_u^{aq}} \tag{14}$$

$$K_{sw} = \frac{c_u^s (1 - c_3 v_3)}{c_u^{aq}} \tag{15}$$

What is important here is the factors $\frac{1 - c_3 v_3}{c_3 v_3}$ and $(1 - c_3 v_3)$ used in deriving Eqs. (14) and (15); note that $\frac{c_u^m}{c_u^{aq}}$ depends not only on the relative affinity of the solute in two subphases but also the relative amounts of the micellar subphase in the solution.[27–29] To extract the relative affinity information only, the subphase amounts must be normalized, which has led to the factors $1 - c_3 v_3$ and $c_3 v_3$ in Eqs. (14) and (15). Combining Eqs. (13)-(15) yields

$$K = \frac{K_{sw}}{1+(K_{mw}-1)v_3 c_3} \tag{16}$$

Substituting Eq. (16) into Eq. (8) yields

$$\Delta G_{u3} - \Delta G_{u1} = \frac{(K_{mw}-1)v_3}{1+(K_{mw}-1)v_3 c_3} \tag{17}$$

Thus, the micelle-water binding constant, K_{mw}, from the stoichiometric model, when used solely as a fitting parameter, can directly yield the KBI difference via Eq. (17), thereby enabling an evaluation of weak, non-specific interactions realistically via statistical thermodynamics. This analysis remains valid even when K_{mw} is negative, i.e. the antibinding state. The app allows the user to input K_{mw} at a chosen surfactant concentration c_2' plus an estimate of v_3 to perform the conversion to ΔG_{u3}, again assuming that the ΔG_{u1} term is negligible. The value shown in Figure 3 (the lower portion of the app in Figure 2) using Armstrong and Stine's own value of K_{mw} is similar (see the mouse read-out at 10.2 mM) to the values calculated via the full theory.

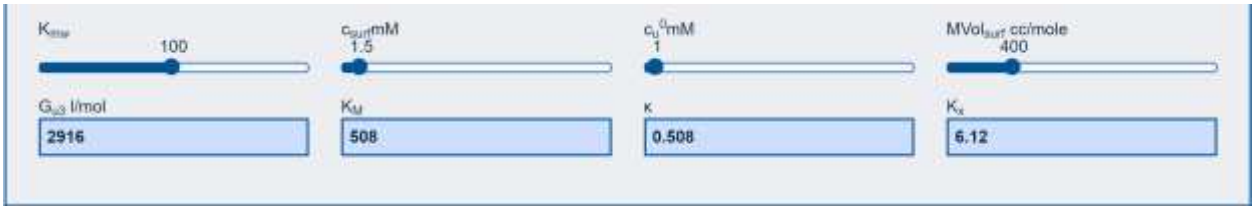

Figure 3. Calculation of solute-micelle KBI (ΔG_{u3}) from empirically-determined K_{mw} values reported in the literature, along with experimentally-determined properties of surfactant and solute [28,29] required for KBI determination. This is part of the MLC app presented in Figure 2.

Not only does Eq. (17) serve as a convenient method for quantifying weak, non-specific interactions through the evaluation of KBIs using the stoichiometric binding model purely as a

fitting model for experimental data; it also clarifies the key differences in molecular picture between stoichiometric binding and statistical thermodynamics. In addition, the important limitation of the stoichiometric approach, which has posed severe difficulties in the understanding of protein denaturation and stability from 1970s to early 2000s,[6,30,31,58,59] has come to the fore.

Firstly, while the negative K_{mw} ("antibinding") in the framework of the binding model signifies the failure and limitation of the model, statistical thermodynamics interprets this as a negative $\Delta G_{u3} - \Delta G_{u1}$, which represents a net depletion of cosolute around the solute as compared to the bulk.[36] Unlike the solvent exchange model, that attempted to incorporate cosolute exclusion within stoichiometric model, statistical thermodynamics is free from assumptions, such as how many water molecules should replace cosolute on solvation shell or the thickness of the solvation shell.[8,36] Cosolute exclusion in general has been well-documented around macromolecules as well as around small solutes,[7–11] and Eq. (17) now makes it possible to quantify micelle exclusion from chromatographic data. However, experimental results on antibinding analysed via our theory yield cosolute exclusion far larger than micelle-solute exclusion volume. This supports the common-sense notion, contrary to the implicit assumption required by MLC theory, that surfactants not only affect the interaction between the solute and the mobile phase but also between the solute and the stationary phase. The evidence from NMR,[60] suggesting the capability of anionic surfactants to bind to some mobile phases, seems to be consistent with this reasoning.

Secondly, the $K_{mw} - 1$ term in Eq. (17) shows that even a weak positive binding constant (for which $K_{mw} < 1$) gives rise to a negative $\Delta G_{u3} - \Delta G_{u1}$, which again signifies the exclusion of micelles. That weakly-binding cosolutes can nevertheless lead to preferential exclusion posed a

serious paradox in the elucidation of protein denaturation and stabilitzation.[6,30,31,58,59] The clue to resolving this paradox, albeit phenomenologically, came from considering the competitive binding of cosolute and water on the shared binding sites on biomolecules.[31] The chromatographic binding model also takes into account both the binding of water and cosolutes, which has given rise to $K_{mw} - 1$. This has at least captured the basic relationship between weak binding and preferential exclusion right, albeit in a primitive model.

The lower portion of the MLC app shown in Figure 3 allows this translation to be made. From the user's K_{mw}, concentration and molar volume of the surfactant, and the concentration of the solute, the G_{u3} and other values are calculated.

The key limitation of this approach is that the values are based on a linear average across concentrations whereas the KBI approach creates values from the real-world, non-linear dependency at each concentration. So, the mining of literature data needs to be done with due regard to this (modest) limitation.

3.3 Comparing MLC to solubility data

One of the major hindrances towards the high-throughput measurements of micellar solubilization comes from the different theoretical models used in analyzing such properties as solubility and chromatographic behavior in the framework of stoichiometric binding models. Since there is now a common theoretical framework for these two, the results from the binding model for chromatography and the partitioning model for solubility can be compared directly. To this end,

here we briefly revisit the two key parameters used in the partitioning model for solubility κ.[48,61,62] The first is the molar solubilization ratio (MSR) κ at a total surfactant concentration c_2', which has been defined as

$$c_u = c_u^0 + \kappa(c_2' - c_2^{cmc'}) \quad (18)$$

where c_u is the solubility of the solute and c_u^0 is that at CMC ($c_2^{cmc\prime}$).[48,61,62] The second is the *molar* micelle-water partition coefficient K_M, which appears in the context of rewriting Eq. (18) as [49]

$$\frac{c_u}{c_u^0} = 1 + \frac{\kappa\left(c_2' - c_2^{cmc'}\right)}{c_u^0} = 1 + K_M\left(c_2' - c_2^{cmc'}\right) \simeq 1 + K_M c_2' \quad (19)$$

where $c_2' \gg c_2^{cmc'}$ is usually the concentration region used for solubilization.[48,61,62] The "partition coefficient" K_M here refers to the partitioning of solute molecules between water and micellar interior, which is a further simplification of the stoichiometric modelling of solute-surfactant complexation.[63] Using Eq. (19) as K in Eq. (7), we obtain[49]

$$G_{u3}^0 - G_{u1}^0 = \frac{mK_M}{1+mK_M c_3} \quad (20)$$

Thus, there is a one-to-one correspondence between the chromatographic micelle-water binding constant and the solubility-based micelle-water partitioning coefficient, as can be seen by comparing Eqs. (17) and (20):

$$mK_M = v_3(K_{mw} - 1) \quad (21)$$

Thus, by treating K_M and K_{mw} as purely fitting parameters, the experimental data in the literature can now be used to yield solute-micelle KBIs both from the solubility and chromatographic measurements. Thus, if the assumptions underlying MLC are correct, solubility and chromatographic behaviour, expressed in terms of K_M and K_{mw}, are complementary to one another in determining the solute-micelle affinities quantified via KBIs.

However, a systematic comparison (necessarily limited by the difficulties of finding values for the same solutes and surfactants) between solubility expressed by mK_M and MLC parameter $v_3(K_{mw} - 1)$ in Figure 4 shows the discrepancy between the two, underscoring our doubts on the basic assumption of MLC (no interaction of the surfactant with the stationary phase) presented in Section 3.2.

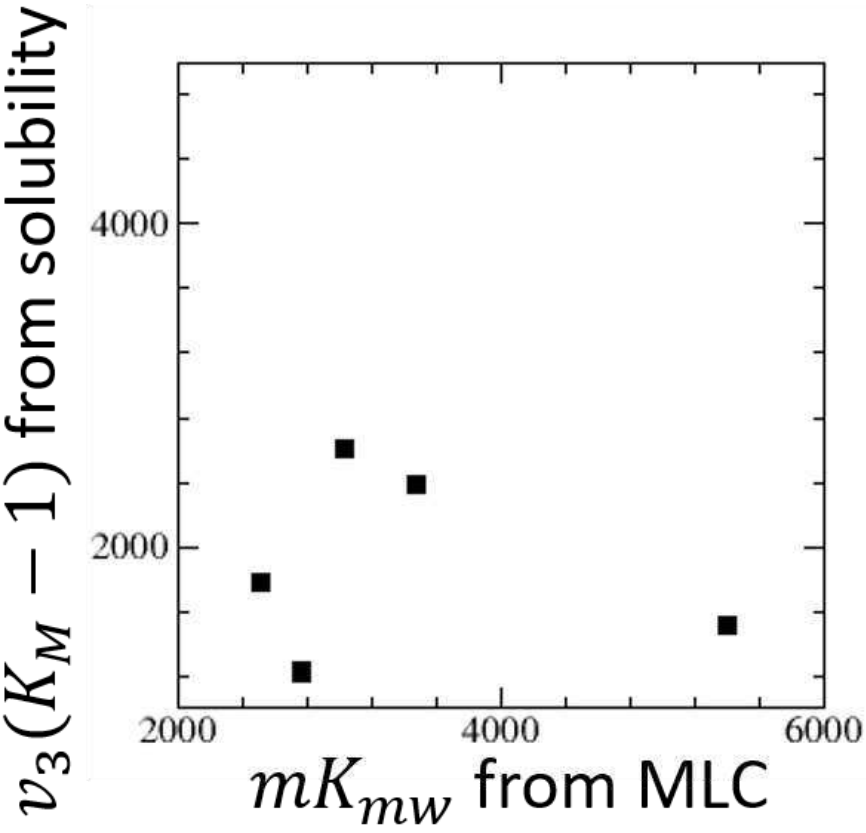

Figure 4. Correlation between micelle-water partition coefficients determined from MLC (M_{mw})[51,64] and solubility (K_M)[65] for substituted phenols in the presence of SDS. A poor correlation raises questions on the basic assumption of MLC.

3.4. A universal statistical thermodynamic measure of chromatographic affinity

We have chosen, in the present paper, the two chromatographic techniques that deal with cosolutes with different degrees of self-association. Until recently, understanding the mechanism of solubilization through the addition of cosolutes had long been hampered severely by the lack of a general theory. Cosolutes have been categorized by the degrees of self-association, from "cosolvents" (whose with weak self-association) to "hydrotropes" (self-associating but not strong enough to form micelles) and micelles.[2,66,67] Each category of solubilization was modelled and explained differently, often posing questions as to which theory to adopt for a given cosolute (e.g., urea as a hydrotrope).[68–70] Our recent universal theory of cosolute solubilization,[4,9] as well as the universal formulation of the cosolute effect in the present paper, forces us to reconsider how we should approach micelle-solute interactions.

The discrepancy is due to the following difference in perspective:

a. incorporation of solute into a micelle;
b. distribution of (non-aggregating) cosolutes around a solute.

Statistical thermodynamics endorses (b), and advocates that micellar solubilization should also be formulated according to (b), thereby establishing a unified approach applicable to all cosolutes regardless of their self-aggregation. Indeed, solubilization is the change of insertion free energy (i.e., Ben-Naim's pseudochemical potential) of a solute molecule,[71] which can be explained by the distribution of cosolutes around a solute molecule.[1,2] These two perspectives are equivalent when one deals with KBIs, due to their symmetry, $G_{ui} = G_{iu}$,[3] which shows that cosolute distribution around a solute is equivalent to solute distribution around a micelle. However, this equivalence breaks down when we explicitly consider the excess coordination number, defined as $N_{ij} =$

$c_j G_{ij}$.[68] This means that the excess coordination number of micelles around a solute, N_{u3} is different from that of solutes around a micelle, N_{3u},[66,69] as can easily be proven from its definition. Most importantly, for sparingly soluble solutes, N_{ui} is useful but N_{iu} is not, because the latter tends to zero at $c_u \to 0$. For convenience, the key N_{ij} values are provided in the apps. We might prefer to think of excess solutes in a micelle, but the thermodynamics tells us that we must get used to thinking of it the other way round.

Thus, we have shown that, for clarity, cosolute distribution around a solute should be adopted as a universal measure of chromatographic affinity.

4. Conclusion

Quantifying non-specific interactions in aqueous solutions is crucial for controlling aqueous solvation and solubility through the change of solvent composition, and chromatography offers a route to high-throughput determination of such interactions. However, the analysis and interpretation of chromatographic data has long been reliant upon the assumption of stoichiometric solute-water and solute-cosolute binding, which has given rise to paradoxes, such as the "antibinding" (negative solute-cosolute binding constant)[27–29] and the indeterminacy of salt-ligand affinity.[18] Such paradoxes have posed a great difficulty, because they are contradictory to the binding models themselves and have shown their serious limitations.

Hence, we propose a rigorous statistical thermodynamic approach, which can capture non-specific interactions, free from stoichiometric binding assumptions. Both solute binding on (or in)

the cosolute and cosolute exclusion from the solute (or, equivalently, solute exclusion from the cosolvent) can be quantified and captured in terms of the Kirkwood-Buff integrals,[33–35] which represent a net excess or depletion of cosolute around the solute as compared to the bulk, based on the molecular distribution function. The chromatographic measurements, regardless of the degree of cosolute self-association, can now be interpreted in a universal theoretical framework that has also been used to rationalize small solute solubility,[4,9,68–70] biomolecular stability, binding, aggregation and gelation.[7,8,36,37] In addition, arbitrary assumptions regarding the stationary phase, or even the neglect thereof, have also been eliminated by statistical thermodynamics. To allow this general approach to be widely used, we have provided two open-source apps which analyse a number of representative datasets from the literature and where users can load and analyse their own datasets.

In view of the decades-long tradition of analysing chromatographic data based on stoichiometry, we have also demonstrated how KBIs can directly be obtained from such an analysis, by treating the binding model merely as a model for data fitting.[39] Such an approach has paved a way towards utilizing the wealth of literature data in a straightforward manner, while eliminating any concern about the problems of the stoichiometric model itself.

Acknowledgements

We are grateful to Dr Alina Koneva and colleagues at St Petersburg State University for providing the raw MLC data accompanying their rather full analysis of MLC effects in the context of NMR and light scattering data with a range of solutes and two surfactants. We thank Dr. Laura Waters

for bringing our attention to the antibinding problem in MLC and Professor Jane Thomas-Oates for continuous encouragements.

References

1 S. E. Friberg, *Curr. Opin. Colloid Interface Sci.*, 1997, **2**, 490–494.

2 W. Kunz, K. Holmberg and T. Zemb, *Curr. Opin. Colloid Interface Sci.*, 2016, **22**, 99–107.

3 S. Abbott, J. J. Booth and S. Shimizu, *Green Chem.*, 2017, **19**, 68–75.

4 S. Shimizu and N. Matubayasi, *Phys. Chem. Chem. Phys.*, 2017, **19**, 23597–23605.

5 S. N. Timasheff, *Adv. Protein Chem.*, 1998, **51**, 355–432.

6 S. N. Timasheff, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 9721–6.

7 S. Shimizu and N. Matubayasi, *J. Phys. Chem. B*, 2014, **118**, 3922–3930.

8 S. Shimizu, R. Stenner and N. Matubayasi, *Food Hydrocoll.*, 2017, **62**, 128–139.

9 S. Shimizu and N. Matubayasi, *J. Phys. Chem. B*, 2014, **118**, 10515–10524.

10 S. Shimizu and N. Matubayasi, *Biophys. Chem.*, 2017, **231**, 111–115.

11 S. Shimizu and N. Matubayasi, *Phys. A Stat. Mech. its Appl.*, 2018, **492**, 1988–1996.

12 J. C. Lee, K. Gekko and S. N. Timasheff, *Methods Enzymol.*, 1979, **61**, 26–49.

13 A. S. Koneva, E. Ritter, Y. A. Anufrikov, A. A. Lezov, A. O. Klestova, N. A. Smirnova, E. A. Safonova and I. Smirnova, *Colloids Surfaces A*, 2018, **538**, 45–55.

14 A. Boichenko, L. Loginova and A. Kulikov, *Методы И Объекты Химического Анализа*, 2007, **2**, 92–116.

15 E. M. Basova, V. M. Ivanov and O. A. Shpigun, in *Encyclopedia of Separation Science, Vol. II*, eds. M. Cooke and C. F. Poole, Academic Press, London, 2000, pp. 726–737.

16 N. Memon, H. I. Shaikh and A. R. Solangi, *Chromatogr. Res. Int.*, 2012, **2012**, 1–6.

17 T. M. Kalyankar, P. D. Kulkarni, S. J. Wadher and S. S. Pekamwar, *J. Appl. Pharm. Sci.*, 2014, **4**, 128–134.

18 I. Slama, C. Ravelet, C. Grosset, A. Ravel, A. Villet, E. Nicolle and E. Peyrin, *Anal. Chem.*, 2002, **74**, 282–287.

19 E. Peyrin, Y. C. Guillaume and C. Guinchard, *Biophys. J.*, 1999, **77**, 1206–12.

20 Y. C. Guillaume, E. Peyrin, A. Villet, A. Nicolas, C. Guinchard, J. Millet and J. F. Robert, *Chromatographia*, 2000, **52**, 753–757.

21 J. Hermansson and I. Hermansson, *J. Chromatogr. A*, 1994, **666**, 181–191.

22 J. Hermansson and A. Grahn, *J. Chromatogr. A*, 1995, **694**, 57–69.

23 S. Allenmark, B. Bomgren and H. Boren, *J. Chromatogr. A*, 1984, **316**, 617–624.

24 G. Örn, K. Lahtonen and H. Jalonen, *J. Chromatogr. A*, 1990, **506**, 627–635.

25 L. A. Sly, D. L. Reynolds and T. A. Walker, *J. Chromatogr. A*, 1993, **641**, 249–255.

26 J. Yang and D. S. Hage, *J. Chromatogr. A*, 1996, **725**, 273–285.

27 D. W. Armstrong and F. Nome, *Anal. Chem.*, 1981, **53**, 1662–1666.

28 D. W. Armstrong and G. Y. Stine, *Anal. Chem.*, 1983, **55**, 2317–2320.

29 D. W. Armstrong and G. Y. Stine, *J. Am. Chem. Soc.*, 1983, **105**, 2962–2964.

30 J. A. Schellman, *Biopolymers*, 1978, **17**, 1305–1322.

31 J. A. Schellman, *Biopolymers*, 1987, **26**, 549–559.

32 J. A. Schellman, *Biophys. Chem.*, 2002, **96**, 91–101.

33 J. G. Kirkwood and F. P. Buff, *J. Chem. Phys.*, 1951, **19**, 774–777.

34 D. G. Hall, *Trans. Faraday Soc.*, 1971, **67**, 2516–2524.

35 A. Ben-Naim, *J. Chem. Phys.*, 1977, **67**, 4884–4890.

36 S. Shimizu, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 1195–1199.

37 S. Shimizu and C. L. Boon, *J. Chem. Phys.*, 2004, **121**, 9147–9155.

38 S. Shimizu, W. M. McLaren and N. Matubayasi, *J. Chem. Phys.*, 2006, **124**, 234905.

39 S. Shimizu, S. Abbott and N. Matubayasi, *Food Funct.*, 2017, **8**, 2999–3009.

40 S. Shimizu and S. Abbott, *J. Phys. Chem. B*, 2016, **120**, 3713–3723.

41 I. W. Wainer, *J. Chromatogr. A*, 1994, **666**, 221–234.

42 I. M. Chaiken, *J. Chromatogr. B Biomed. Sci. Appl.*, 1986, **376**, 5–9.

43 A. J. Muller and P. W. Carr, *J. Chromatogr. A*, 1984, **284**, 33–51.

44 J. E. S. J. Reid, A. J. Walker and S. Shimizu, *Phys. Chem. Chem. Phys.*, 2015, **17**, 14710–14718.

45 J. E. S. J. Reid, R. J. Gammons, J. M. Slattery, A. J. Walker and S. Shimizu, *J. Phys. Chem. B*, 2017, acs.jpcb.6b10562.

46 A. A. Dar, G. M. Rather and A. R. Das, *J. Phys. Chem. B*, 2007, **111**, 3122–3132.

47 P. A. Bhat, G. M. Rather and A. A. Dar, *J. Phys. Chem. B*, 2009, **113**, 997–1006.

48 C. De Oliveira, R. Yagui, A. Lineu Prestes, C. O. Rangel-Yagui, A. Pessoa-Jr and L. C. Tavares, *J Pharm Pharm. Sci*, 2005, **8**, 147–163.

49 N. Matubayasi, K. K. Liang and M. Nakahara, *J. Chem. Phys.*, 2006, **124**, 154908.

50 Y. Shahzad, University of Huddersfield, 2013.

51 M. L. Marina and M. A. García, *J. Chromatogr. A*, 1997, **780**, 103–116.

52 M. L. Marina and M. A. Garcia, in *Handbook of methods and instrumentation in separation*

science, Vol II, eds. I. D. Wilson and C. F. Poole, Elsevier, Amsterdam, 2000, pp. 726–737.

53 J. A. Schellman, *Biopolymers*, 1975, **14**, 999–1018.

54 K. C. Aune and C. Tanford, *Biochemistry*, 1969, **8**, 4579–4585.

55 K. C. Aune and C. Tanford, *Biochemistry*, 1969, **8**, 4586–4590.

56 E. F. Casassa and H. Eisenberg, *Adv. Protein Chem.*, 1964, **19**, 287–395.

57 I. M. Klotz, *Ligand-receptor energetics : a guide for the perplexed*, Wiley, New York, 1997.

58 S. N. Timasheff, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 7363–7367.

59 V. A. Parsegian, R. P. Rand and D. C. Rau, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 3987–3992.

60 B. K. Lavine, S. Hendayana, Y. He and W. T. Cooper, *J. Colloid Interface Sci.*, 1996, **179**, 341–349.

61 D. Attwood and A. T. Florence, in *Surfactant Systems*, Springer Netherlands, Dordrecht, 1983, pp. 229–292.

62 F. A. Alvarez-Núñez and S. H. Yalkowsky, *Int. J. Pharm.*, 2000, **200**, 217–222.

63 E. H. K. Shinoda, *J. Phys. Chem.*, 1961, **66**, 577–582.

64 J. P. Foley, *Anal. Chim. Acta*, 1990, **231**, 237–247.

65 Y. P. Saraf and S. S. Bhagwat, *Sep. Technol.*, 1995, **5**, 207–212.

66 S. E. Friberg, *Curr. Opin. Colloid Interface Sci.*, 1997, **2**, 490–494.

67 P. Bauduin, A. Renoncourt, A. Kopf, D. Touraud and W. Kunz, *Langmuir*, 2005, **21**, 6769–6775.

68 J. J. Booth, S. Abbott and S. Shimizu, *J. Phys. Chem. B*, 2012, **116**, 14915–14921.

69 S. Shimizu, J. J. Booth and S. Abbott, *Phys. Chem. Chem. Phys.*, 2013, **15**, 20625–20632.

70 J. J. Booth, M. Omar, S. Abbott and S. Shimizu, *Phys Chem Chem Phys*, 2015, **17**, 8028–8037.

71 A. Ben-Naim, *Molecular theory of solutions*, Oxford University Press, Oxford, 2006.

Graphical abstract